

# ANALISI DELLE FREQUENZE DI ACQUISTO NELL' INDAGINE SUI CONSUMI DELLE FAMIGLIE

Ignazio Drudi\*

Rapporto di ricerca n.8

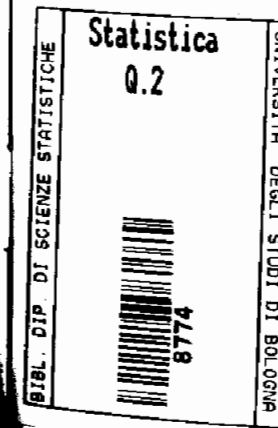
CON PRI - La misura dei consumi privati

I lavori raccolti in questa collana hanno avuto origine nell'ambito del progetto di ricerca dell'ISTAT «Le statistiche dei consumi privati nel sistema statistico nazionale» e del progetto di ricerca MURST 40% «La misura dei consumi privati: uno studio sull'accuratezza, coerenza e qualità dei dati».

Al progetto di ricerca hanno partecipato i ricercatori dell'ISTAT e dei seguenti Dipartimenti e Istituti universitari:

- Dipartimento di Scienze Statistiche, Bologna
- Dipartimento Statistico, Firenze
- Dipartimento di Scienze Statistiche, Perugia
- Dipartimento di Contabilità Nazionale, Roma
- Istituto di Statistica e Matematica, Istituto Universitario Navale, Napoli
- Istituto di Statistica, Messina.

\* Dipartimento di Scienze Statistiche - Università di Bologna



Dipartimento di Scienze Statistiche "Paolo Fortunati"  
dell' Università degli Studi di Bologna  
Novembre 1992

## INDICE

1. Le frequenze di spesa e la distinzione tra spesa e consumo	p. 5
2. Le frequenze di spesa e l'efficienza delle stime	p. 9
3. Le frequenze di spesa e la numerosità campionaria	p. 19
4. Le frequenze di spesa e i modelli di comportamento individuali	p. 22
4.1 La probabilità di spesa nel modello Tobit	p. 22
4.2 Il comportamento di spesa come processo stocastico	p. 24
5. Le frequenze di spesa e la ricerca di tipologie di consumatori	p. 25
6. Considerazioni conclusive	p. 32
<i>Note</i>	p. 33
<i>Riferimenti bibliografici</i>	p. 34
<i>Appendice</i>	p. 37

## 1. Le frequenze di spesa e la distinzione tra spesa e consumo

Nell'enunciare le principali finalità dell'indagine sui bilanci di famiglia (d'ora in avanti BF), l'ISTAT sottolinea che "l'indagine sui consumi delle famiglie ha lo scopo primario di rilevare la struttura e il livello dei consumi secondo modalità di carattere economico-sociale e territoriale. La rilevazione viene svolta con il "Metodo della spesa" che consiste nel rilevare tutti i beni e servizi acquistati o comunque consumati dalle famiglie per le proprie necessità" (ISTAT, 1990).

Come è noto<sup>1</sup>, l'indagine è di tipo campionario e prevede il coinvolgimento di circa 3000 nuclei familiari ogni mese ai quali viene richiesto, in primo luogo, di registrare, giorno per giorno, su un apposito libretto tutte le spese effettuate nel corso di una decade. In secondo luogo, nel corso di una intervista, le famiglie vengono stimolate a ricostruire tutti gli acquisti di beni non alimentari effettuati nel corso del mese, e a ricordare le spese per beni durevoli sostenute nell'arco dei tre mesi precedenti.

L'oggetto del processo di misura è dunque l'ammontare delle spese sostenute dal nucleo familiare in un determinato intervallo di tempo. Tale modo di procedere è comune a tutti i paesi in cui vengono svolte indagini sui consumi familiari e, implicitamente, identifica nella spesa la manifestazione fenomenica misurabile del consumo.

Il ricorso al metodo della spesa nella misura dei consumi familiari è evidentemente una scelta obbligata, poiché questa è l'unica metodologia che consente di rilevare simultaneamente ammontare dell'acquisto e caratteristiche dell'acquirente, la cui conoscenza costituisce la premessa per la individuazione di comportamenti individuali, ed è, quindi, indispensabile per conseguire l'obiettivo di cogliere modalità e livelli di consumo di sottoinsiemi di famiglie. Tuttavia tale scelta non è senza conseguenze, o per meglio dire, richiede l'esplicitazione della relazione che lega la grandezza che si vuole misurare (il consumo) alla sua manifestazione fenomenica (la spesa). Una trattazione approfondita di questo problema<sup>2</sup> va, evidentemente, al di là degli obiettivi di questo lavoro. In questa sede ci limiteremo a richiamare un aspetto che caratterizza la relazione fra spesa e consumo: quello relativo alla cadenza temporale dell'evento che viene osservato, cioè l'acquisto di un bene o servizio.

Nella sua accezione più generale<sup>3</sup>, il consumo è definito come il processo attraverso il quale beni e servizi, originati da un processo di produzione, vengono utilizzati dagli individui per il soddisfacimento dei propri bisogni.

D'altra parte la spesa è la quantità di potere d'acquisto, cioè di reddito, che, in un istante o in un dato periodo di tempo, ciascun individuo indirizza verso beni e servizi in grado di soddisfare i suoi bisogni. E' piuttosto agevole mostrare<sup>4</sup> che in una prospettiva temporale indefinita i due fenomeni tendono a coincidere, nel senso che qualsiasi acquisto verrà impiegato per soddisfare qualche bisogno (se non altro il desiderio di possesso), così come, compatibilmente con i vincoli di reddito, ogni bisogno darà origine ad un acquisto.

Tuttavia, ed è questo il punto cruciale dell'argomentazione, il processo di misura non può astrarre dalla dimensione temporale: esso è confinato in un dominio temporale definito e limitato e, in un siffatto dominio, la relazione che lega spesa e consumo diviene tutt'altro che scontata. In altri termini, si può sostenere che, da un lato, le principali teorie economiche del consumatore fanno riferimento a comportamenti di consumo medi di lungo periodo, o addirittura collocati in una prospettiva a-temporale; dall'altro lato, le misure che sono disponibili riguardano una variabile che non solo è diversa per definizione, ma, nel breve periodo, può diventare un indicatore assai debole del consumo<sup>5</sup>. Infatti, solo a lungo termine, la decisione di consumare e quella di acquistare saranno influenzate dai medesimi fattori e avverranno nelle stesse condizioni; al contrario nel breve periodo, accanto alla determinazione di cosa e quanto consumare, assume importanza cruciale anche la valutazione del quando (o meglio, con quale frequenza) effettuare gli acquisti. L'obiettivo principale di questo lavoro è proprio il tentativo di valutare in che modo la frequenza con cui avvengono gli acquisti influenza il processo di misura.

Il processo decisionale che porta la famiglia a determinare la cadenza dei suoi acquisti è la risultante di una pluralità di fattori che non necessariamente svolgono un ruolo di rilievo nelle scelte che riguardano le modalità di soddisfacimento dei bisogni, cioè con l'attività di consumo. Il più importante di questi fattori è costituito dalle caratteristiche merceologiche dei beni, in larga parte legate al livello e alla diffusione della tecnologia di produzione e conservazione. Alcuni beni, infatti, sono in grado di soddisfare più volte nel tempo il medesimo bisogno (beni durevoli), altri non perdono la loro capacità di soddisfare bisogni con il passare del tempo (beni conservabili), per altri ancora il solo possesso rappresenta il bisogno da soddisfare (beni artistici e beni-simbolo).

Un secondo fattore, assai rilevante nella determinazione della frequenza degli acquisti, è costituito dal livello di complessità organizzativa e logistica

dell'approvvigionamento soprattutto in ambito urbano. Il tempo richiesto per gli spostamenti, l'innalzamento della quota di donne occupate, la crescente dotazione di strumenti di conservazione domestica dei cibi, la diffusione della grande distribuzione commerciale, spingono con forza il nucleo familiare a concentrare gli acquisti in determinati giorni della settimana, o in dati periodi dell'anno. Anche per una certa parte dei beni alimentari, è oggi possibile che il momento dell'acquisto sia svincolato da quello del consumo.

In sintesi, un primo elenco, sicuramente parziale di variabili determinanti per la frequenza degli acquisti, dovrebbe contemplare: la tipologia delle merci da acquistare, la organizzazione familiare, la tipologia dei servizi commerciali e le caratteristiche del luogo di residenza del nucleo familiare, i gusti del singolo consumatore, la dotazione di beni strumentali in possesso della famiglia, i modelli di comportamento dominanti, il tipo di attività svolta dai singoli membri.

La relazione che lega consumo e spesa è dunque assai complessa: alcuni dei fattori elencati influenzano unicamente la decisione d'acquisto, altri hanno rilevanza sia nel comportamento di spesa che in quello di consumo, altri toccano le due variabili con intensità e direzioni diverse. Tanto che sarebbe possibile ipotizzare una combinazione di fattori tale che due nuclei familiari con comportamento di consumo assai simili possano mostrare abitudini di acquisto del tutto diverse, e viceversa.

Le argomentazioni fin qui esposte mettono in luce uno dei problemi principali dell'indagine sui consumi, infatti, la frequenza con cui si verifica l'evento acquisto è cruciale per il processo di misura. Abbiamo già sottolineato che l'indagine BF rileva acquisti su un periodo di tempo determinato (10, 30 o 90 giorni, a seconda dei beni). A prescindere dalla accuratezza della registrazione, quanto più breve è il periodo di riferimento tanto maggiore è la probabilità che alcune spese non siano osservate. Anche quando la spesa interessa un bene o un gruppo di beni il cui consumo è consistente, la rarità dell'evento di acquisto o la "durabilità" di un bene possono determinare nel periodo di osservazione una registrazione di spesa nulla.

Una delle manifestazioni più evidenti della discrasia che abbiamo illustrato è la forte variabilità con cui si manifesta l'evento di acquisto al variare del tipo di bene e del tipo di famiglia<sup>6</sup>. L'esame della tav. 1 raccoglie una evidenza empirica eloquente circa la variabilità legata alle caratteristiche dei diversi beni: nella tavola sono riportate le frequenze delle registrazioni di spesa per ciascuna delle 178 voci considerate nell'indagine BF con riferimento al IV°

trimestre del 1989. Il 99% delle famiglie ha, almeno una volta, acquistato pane nella decade in cui ha partecipato all'indagine; analogamente percentuali molto alte hanno voci il cui consumo è assai frequente, come alcuni beni alimentari (carne 91%, latte 93%, pasta 90%) o come quelle che risultano essere spese, per così dire, "obbligatorie" (energia elettrica 99%, spese per telefono 84%). Di contro, più dell' 80% delle 178 voci mostra una percentuale inferiore al 50%: in particolare le spese per beni durevoli interessano dall'1 al 10% delle famiglie, la spesa per mobili si colloca fra il 10 e il 20%, e quella per l'abbigliamento si attesta attorno al 30% delle famiglie.

Il problema della presenza di diverse frequenze nell'acquisto dei beni viene solo marginalmente preso in considerazione nell'indagine BF, in questa chiave, infatti, può essere interpretata la coesistenza di diversi riferimenti temporali per la registrazione delle spese (trimestrale per i beni durevoli, mensile per i semidurevoli, decedele per gli alimentari e i correnti) nonché la rilevazione di informazioni circa il possesso di beni durevoli. Tuttavia queste diverse strategie influenzano unicamente le modalità di rilevazione del fenomeno e non hanno effetto sulla definizione e la costruzione degli stimatori. Infatti, per quanto concerne le procedure di stima, le informazioni desunte nei tre riferimenti temporali vengono rese confrontabili assumendo l'ipotesi implicita di comportamento omogeneo delle famiglie nel tempo. L'operazione che consiste nel riportare al mese delle informazioni decedali, calcolata triplicando la spesa decedale per ottenere il dato mensile, e di quelle riferite al trimestre, calcolata dividendo per tre le spese per durevoli, risponde ad un'idea di sostanziale uniformità nei comportamenti che non appare in alcun modo giustificata. Una considerazione analoga può essere riproposta anche per l'avvicendamento di campioni mensili indipendenti nel trimestre.

Al di là di questi accorgimenti organizzativi e di calcolo, il punto della questione è che né il piano di campionamento né gli stimatori ad esso collegati sono progettati per affrontare esplicitamente il problema delle diverse frequenze di spesa. Il problema non è di poco conto dal momento che gli effetti della presenza di spese rare si manifesta sia nelle stime di livello, cioè nella misura dell'ammontare della spesa di tutte le famiglie per ciascun bene, sia nelle stime dei valori medi familiari. Pertanto, il problema di una corretta rilevazione delle spese rare si presenta sia quando si pone mano al confronto fra i dati dell'indagine BF e i dati desunti da altre fonti, in particolare quelli ottenuti nell'ambito della Contabilità Nazionale, sia quando ci si occupa delle

stime pubblicate che traggono origine dall'indagine, che riguardano valori medi di spesa mensile, familiare o pro-capite.

Questa obiettiva difficoltà del processo di rilevazione nel tenere conto della esistenza di spese rare si riverbera ovviamente anche sulle possibilità di utilizzo dei dati ai fini dell'analisi economica. Come verrà mostrato in seguito tutte le principali tecniche statistiche usualmente utilizzate per studiare il fenomeno risentono pesantemente della presenza di zeri di spesa<sup>7</sup>, all'interno dei quali non è possibile distinguere fra effettiva assenza di consumo e inadeguatezza dell'osservazione empirica.

Il lavoro qui presentato tratta di questa duplice difficoltà, che concerne il processo di misura e i metodi di analisi statistica. I primi due paragrafi hanno lo scopo di evidenziare l'effetto della rarità di alcuni acquisti sulla strategia di campionamento, con particolare riferimento alla variabilità degli stimatori e alla determinazione della numerosità campionaria ottimale. Nei quattro paragrafi seguenti vengono esaminati gli effetti delle frequenze di spesa sulla stima econometrica di modelli individuali, sulla definizione di modelli stocastici di spesa, sulle metodologie di analisi fattoriale, e, infine, sulle tecniche di raggruppamento.

Più precisamente, nei primi paragrafi si tenterà di dimostrare che per molti beni di cui l'indagine BF rileva le spese, il piano di campionamento si rivela inadeguato per la stima efficiente delle frequenze e dell'ammontare della spesa. Tale carenza ha evidentemente effetti importanti sulla stima dell'ammontare dei consumi, ma ancora più rilevante è l'impatto di una stima non efficiente sull'utilizzo che si può fare dei dati forniti dall'indagine. Per affrontare questo tema, nei successivi paragrafi, si passerà ad analizzare gli effetti delle frequenze di spesa sul versante della *relevance* dell'indagine, focalizzando l'attenzione sul ruolo che esse giocano nelle principali tecniche statistiche di norma utilizzate per l'analisi del fenomeno consumi.

Ciascun argomento è affrontato con specifico riferimento alla rilevazione BF così come è condotta attualmente. In particolare i dati utilizzati per tutte le elaborazioni riportate sono derivati dalla indagine del IV° trimestre 1989.

## 2. Le frequenze di spesa e l'efficienza delle stime

Quando l'obiettivo del piano di campionamento è costituito dalla stima di livelli, cioè dei totali della spesa per ciascun bene, in assenza di qualunque

informazione *a priori* circa la frazione degli acquirenti nell'universo, alla variabilità dello stimatore della spesa media (familiare o pro-capite) si aggiunge quella dovuta alla stima della proporzione degli acquirenti (Cochran, 1977).

Se si adotta, in via di prima approssimazione, l'ipotesi di un campionamento casuale semplice, lo stimatore comunemente proposto per la stima del totale di una variabile d'interesse (ad esempio la spesa totale per il bene i-esimo) è:

$$y_i^* = \frac{N}{n} \sum_{h=1}^n y_{ih} \quad (1)$$

dove:

$y_i^*$  = stima della spesa totale per il bene i-esimo

$N$  = numerosità universo

$n$  = numerosità campione

$y_{ih}$  = spesa totale della famiglia  $h$  per il bene i-esimo.

Nel caso in esame, che prevede la presenza nell'universo e nel campione di unità con spesa pari a zero (non acquirenti) e di unità con spesa diversa da zero (acquirenti), è evidente che l'informazione circa la stima del fenomeno è ricavata totalmente dalle osservazioni nel dominio degli acquirenti. Nel caso di stima del totale, e, come vedremo, anche del valor medio, il sottoinsieme dei non acquirenti gioca un ruolo solo in termini di numerosità, o meglio, di peso relativo. E', infatti, immediato ricavare:

$$\sum_{h=1}^n y_{ih} = \sum_{k=1}^{n_a} y_{ik} \quad (2)$$

e

$$\bar{y}_i = \frac{1}{n} \sum_{h=1}^n y_{ih} = \frac{1}{n} \sum_{k=1}^{n_a} y_{ik} = \frac{n_a}{n} \bar{y}_{ai} \quad (3)$$

dove le sommatorie rispetto al deponente  $k$  sono estese al solo dominio degli acquirenti ed è:

$\bar{y}_i$  = valore medio familiare della spesa per il bene i-esimo nel campione

$\bar{y}_{ai}$  = valore medio familiare della spesa per il bene i-esimo nel campione per le sole famiglie acquirenti

$n_a$  = numerosità campionaria delle famiglie acquirenti il bene i-esimo.

Pertanto lo stimatore di livello può essere ridefinito in base ai soli valori campionari rilevati sul sottoinsieme degli acquirenti:

$$y_i^* = N \bar{y}_i = N \left( \frac{n_a}{n} \right) \bar{y}_{ai} \quad (4)$$

Si dimostra che lo stimatore è corretto ed ha varianza, se la frazione di campionamento ( $n/N$ ) è trascurabile, pari a:

$$S^2(y_i^*) = \frac{N^2 s^2}{\sqrt{n}} \quad (5)$$

dove  $s$  è lo scarto quadratico medio campionario calcolato su tutto l'insieme di famiglie compresi i non acquirenti.

Tuttavia è agevole mostrare che la varianza complessiva è funzione del valor medio della spesa degli acquirenti, della varianza nel dominio degli acquirenti e del peso relativo del sottoinsieme degli acquirenti rispetto alla numerosità campionaria totale.

Infatti, dato un collettivo  $C$  di famiglie di numerosità  $n$ , nel quale venga rilevata la spesa ( $y_i$ ) per l'acquisto di un certo bene, definiamo:

$A$  = insieme di famiglie che hanno acquistato il bene

$B$  = insieme di famiglie che non hanno acquistato il bene

$n_a$  = numerosità di  $A$

$n_b$  = numerosità di  $B$

$y_i$  = ammontare della spesa della famiglia i-esima



Si avrà:

$$A \cup B = C$$

$$n_a + n_b = n$$

$$y_i = \begin{cases} y_{ia} & i \in A \\ 0 & i \in B \end{cases}$$

Poniamo inoltre:

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} \quad (6)$$

$$\bar{y}_a = \frac{\sum_{i=1}^{n_a} y_{ia}}{n_a} \quad (7)$$

avremo:

$$Var_t = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2}{n}$$

dove  $Var_t$  è la varianza di spesa calcolata nell'intero collettivo;

$$Var_a = \frac{\sum_{i=1}^{n_a} (Y_i - \bar{Y})^2}{n_a}$$

dove  $Var_a$  è la varianza calcolata nel collettivo degli acquirenti

Siamo interessati a valutare la quantità:

$$(Var_t - Var_a)$$

cioè la differenza tra la varianza calcolata in tutto il collettivo o nel solo collettivo degli acquirenti.

Si avrà:

$$(Var_t - Var_a) = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n} - \frac{\sum_{i=1}^{n_a} (y_{ia} - \bar{y}_a)^2}{n_a} = \quad (8)$$

$$= \frac{\sum_{i=1}^n y_i^2}{n} - \bar{y}^2 - \frac{\sum_{i=1}^{n_a} y_{ia}^2}{n_a} + \bar{y}_a^2 \quad (9)$$

Poiché è:

$$\bar{y}^2 = \frac{n_a^2}{n^2} \bar{y}_a^2 \quad \text{e} \quad \sum_{i=1}^n y_i^2 = \sum_{i=1}^{n_a} y_{ia}^2$$

avremo allora, sostituendo e raccogliendo i fattori comuni nella (9):

$$\sum_{i=1}^n y_i^2 \left( \frac{1}{n} - \frac{1}{n_a} \right) - y_a^2 \left( \frac{n_a^2}{n^2} - 1 \right) \quad (10)$$

ma la quantità nell'ultima parentesi della (10) è:

$$\left( \frac{n_a^2}{n^2} - 1 \right) = \left( \frac{n_a}{n} - 1 \right) \left( \frac{n_a}{n} + 1 \right) = \left( -\frac{n_b}{n} \right) \left( 2 - \frac{n_a}{n} \right) \quad (11)$$

mentre all'interno della prima parentesi della (10) avremo:

$$\frac{1}{n} - \frac{1}{n_a} = \frac{n_a - n}{nn_a} = \frac{n - n_b - n}{nn_a} = \left( -\frac{1}{n_a} \right) \left( -\frac{n_b}{n} \right) \quad (12)$$

quindi, sostituendo nella formula (10), si ottiene:

$$\sum_{i=1}^n y_i^2 \left( \frac{1}{n_a} \right) \left( -\frac{n_b}{n} \right) - y_i^2 \left( -\frac{n_b}{n} \right) \left( 2 - \frac{n_b}{n} \right) = \quad (13)$$

$$= -\frac{n_b}{n} \left( \frac{\sum_{i=1}^n y_i^2}{n_a} - 2\bar{y}_a^2 + \frac{n_b}{n} y_a^2 \right) = \quad (14)$$

$$= -\frac{n_b}{n} \left( \left( \frac{\sum_{i=1}^n y_i^2}{n_a} - \bar{y}_a^2 \right) + \left( \bar{y}_a^2 \left( \frac{n_b}{n} - 1 \right) \right) \right) \quad (15)$$

Pertanto:

$$(Var_i - Var_a) = -\frac{n_b}{n} \left( Var_a - y_a^2 \frac{n_a}{n} \right) \quad (16)$$

quindi:

$$Var_i = Var_a - \frac{n_b}{n} Var_a + \bar{y}_a^2 \frac{n_a}{n} \frac{n_b}{n} = \quad (17)$$

$$= Var_a \left( 1 - \frac{n_b}{n} \right) + \bar{y}_a^2 \frac{n_a}{n} \frac{n_b}{n} = \quad (18)$$

$$= \frac{n_a}{n} \left( Var_a + \bar{y}_a^2 \frac{n_b}{n} \right) \quad (19)$$

In particolare, avremo  $Var_i = Var_a$  :

i) se  $n_a = n$  cioè quando tutte le famiglie del collettivo acquistano il bene in oggetto

ii) se, vedi (17):

$$-\frac{n_b}{n} Var_a + \bar{y}_a^2 \frac{n_a}{n} \frac{n_b}{n} = 0 \quad (20)$$

dividendo per  $n_b / n$  si ottiene:

$$-Var_a + \bar{y}_a^2 \frac{n_a}{n} = 0 \quad (21)$$

da cui:

$$\frac{n_a}{n} = \frac{Var_a}{\bar{y}_a^2} = (CV_a)^2 \quad (22)$$

cioè se la frazione degli acquirenti è uguale al quadrato del coefficiente di variazione calcolato nel sub- collettivo A.

La (19) mostra chiaramente come la varianza totale della spesa sia funzione della varianza del sottoinsieme A delle famiglie acquirenti, della frazione degli acquirenti e della spesa media degli acquirenti. Quindi è il comportamento delle famiglie acquirenti, attraverso l'ammontare e la frequenza della spesa, a determinare la varianza dello stimatore.

Le conclusioni sono indubbiamente di notevole impatto al fine di valutare la strategia di campionamento dell'indagine BF, e confermano che non esistono giustificazioni di ordine teorico per un piano campionario che rileva l'ammontare della spesa per i diversi beni senza alcuna attenzione alla frequenza del loro acquisto.

Le considerazioni fin qui esposte, relative alla variabilità dello stimatore di livello nel caso di campionamento casuale semplice, possono agevolmente essere estese ad un piano di campionamento stratificato e allo stimatore della media campionaria. In particolare si dimostra (Cochran, 1977) che nel caso di



campionamento stratificato lo stimatore media campionaria nel dominio degli acquirenti assume la forma di uno stimatore rapporto.

In particolare ponendo:

$N_{ha}$  = numerosità degli acquirenti nello strato  $h$  dell'universo

$N_{hb} = N_h - N_{ha}$  = numerosità dei non acquirenti nello strato  $h$  dell'universo

$n_{ha}$  = numerosità degli acquirenti nello strato  $h$  del campione

$n_{hb} = n_h - n_{ha}$  = numerosità dei non acquirenti nello strato  $h$  del campione

si avrà:

$$\bar{y}_{ha} = \frac{\sum_{i=1}^{n_{ha}} y_{ai}}{n_{ha}}$$

media delle spese nel dominio degli acquirenti nello strato  $h$  del campione

$$\bar{Y}_{ha} = \frac{\sum_{i=1}^{N_{ha}} Y_{ai}}{N_{ha}}$$

media delle spese nel dominio degli acquirenti nello strato  $h$  nell'universo.

Gli stimatori del totale e del valor medio della spesa della popolazione nel dominio degli acquirenti sono dati da:

$$Y^* = \sum_h N_{ha} \bar{y}_{ha} \quad \text{e} \quad \bar{Y}^* = \frac{Y^*}{\sum_h N_{ha}}$$

Ovviamente nel nostro caso, le numerosità degli acquirenti in ciascuno strato sono incognite, vanno quindi stimate a partire dai dati campionari. Così facendo, non solo i livelli medi campionari vanno considerati come variabili casuali, ma anche le numerosità, cioè le proporzioni di acquirenti assumono la forma di variabili casuali; pertanto lo stimatore della media assume la forma di uno stimatore rapporto che utilizza come stima del numero degli acquirenti

nell'universo ( $N_a^*$ ) e in ciascuno strato ( $N_{ha}^*$ ):

$$N_a^* = \sum_h N_{ha}^* = \sum_h \frac{N_h}{n_h} n_{ha} \quad (23)$$

In questo caso lo stimatore usuale per campioni stratificati è adeguato allo scopo, in particolare si mantiene la proprietà di correttezza

$$\bar{Y}_a^* = \frac{\sum_h \left( \frac{N_h}{n_h} \sum_{i=1}^{n_h} y_{iha} \right)}{\sum_h \frac{N_h}{n_h} n_{ha}} \quad (24)$$

ma la sua efficienza diminuisce e come era logico attendersi dipende dalla frazione delle famiglie acquirenti. In particolare, ponendo  $f_h = n_{ha} / N_h$ , la variabilità dello stimatore è:

$$V(\bar{Y}_a^*) = \frac{1}{N_a^{*2}} \sum_h \frac{N_h^2 (1 - f_h)}{n_h (n_h - 1)} \left[ \sum_{i=1}^{n_{ha}} (y_{iha} - \bar{y}_{ha})^2 + n_{ha} \left( 1 - \frac{n_{ha}}{n_h} \right) (\bar{y}_{ha} - \bar{y}_a)^2 \right] \quad (25)$$

In particolare, il secondo addendo entro la parentesi quadra nella (25) mostra con chiarezza la relazione che lega precisione della stima e ampiezza del sottoinsieme di famiglie acquirenti. Più precisamente esso rappresenta il contributo della varianza tra gli strati alla varianza dello stimatore che, come si può notare, dipende dalla frazione di acquirenti ( $n_{ha}/n_h$ ). Tale contributo, quindi, costituisce una misura della quota di varianza dovuta alla presenza/assenza di acquirenti. In particolare si nota che si annulla se vale la relazione:

$$\left( 1 - \frac{n_{ha}}{n_h} \right) = \frac{n_{hb}}{n_a} \quad \forall h$$

cioè se la frazione dei non acquirenti è pari a 0 in tutti gli strati.

La tavola 2 riporta i valori delle stime e dell'errore relativo calcolato per i dati del campione BF del IV trimestre '89 e la percentuale di varianza dovuta alla presenza di non acquirenti. Va in primo luogo osservato che solo il 17% dei beni presenta un coefficiente di variazione inferiore a 2.5 punti percentuali (e quindi un errore relativo dello stimatore inferiore al 5%, al 95% di probabilità); mentre, all'estremo opposto si colloca circa un quinto delle categorie di spesa per le quali l'errore relativo è addirittura superiore al 30%. Fra questi due estremi la distribuzione è pressoché uniforme: il 24% dei beni si colloca fra 2.5% e 5%, il 20% fra 5% e 10% e, infine, il 19% fra 10% e 15%.

In sintesi, la stima della spesa media della grande maggioranza dei beni rilevati mostra una variabilità molto pronunciata, in molti casi di portata tale da pregiudicare l'affidabilità della stima. L'affermazione precedente acquista un rilievo ancora maggiore se si considera che è presente una marcata correlazione inversa fra frequenza di spesa e ammontare della stessa. In altri termini, le voci più frequenti, e quindi con minore variabilità, risultano essere le meno rilevanti nella determinazione della spesa totale familiare. Ciò fa sì che, ad esempio, la determinazione del 37% dell'ammontare complessivo della spesa sia fondato su stime per singolo bene che, al minimo, hanno un errore relativo pari al 30%.

Va precisato che il calcolo è stato condotto a livello di disaggregazione più analitico possibile e che successive aggregazioni si compongono secondo una legge additiva (nel senso che è sufficiente un acquisto in una qualsiasi voce dell'aggregato per aumentare la frequenza d'acquisto), tuttavia il comportamento dell'errore appare fortemente caratterizzato e soprattutto le tipologie di beni sono facilmente individuabili. Ad esempio le voci relative a beni durevoli mostrano sistematicamente errori di stima molto alti, che non saranno influenzati da nessuna operazione di aggregazione. Non è, inoltre, difficile notare come le voci più critiche siano le stesse per le quali emergono significative sottostime rispetto alla CN<sup>8</sup>. Anche in questo caso appare difficilmente sostenibile la scelta di non considerare esplicitamente il problema delle frequenze nella definizione della strategia campionaria e soprattutto nella costruzione degli stimatori.

D'altra parte, la relazione che lega frequenza di spesa ed errore di stima è del tutto evidente se si osservano le figg. 1 e 2. In ambedue i grafici è riportata in ascissa la percentuale di famiglie che nel trimestre in esame hanno registrato almeno una spesa, mentre in ordinata sono rappresentati rispettivamente l'ammontare dell'errore relativo dello stimatore (Fig.1) e il contributo alla varianza dello stimatore determinato dalla percentuale di acquirenti (Fig.2);

ciascun punto indicato sui grafici rappresenta una voce di spesa. Emerge con chiarezza la relazione inversa che lega le due variabili, nonché la rapidità con cui decade l'efficienza dello stimatore in presenza di spese rare. In particolare si osserva che per quanto concerne la frequenza di famiglie acquirenti, per un valore compreso tra il 5 e il 10% costituisce una soglia critica al di sotto della quale il coefficiente di variazione dello stimatore presenta livelli che superano il 10% (e quindi un intervallo di confidenza per la stima pari a  $+o-20\%$ ). In corrispondenza della stessa soglia il contributo alla varianza della proporzione dei non acquirenti si fa piuttosto elevato superando il 30% della varianza totale fino a determinare quasi totalmente la varianza della stima.

### 3. Le frequenze di spesa e la numerosità campionaria

Come risulta dalle considerazioni svolte nel paragrafo precedente, la stima del totale e del valor medio delle spese familiari richiede implicitamente o esplicitamente la stima della proporzione di acquirenti e della media delle spese nel sottoinsieme di famiglie che acquistano il bene in esame. Per quanto concerne il primo dei due parametri, il problema è ulteriormente complicato dal fatto che, come risulta evidente dall'esame delle frequenze campionarie, alcuni acquisti si presentano come eventi rari. Com'è noto<sup>9</sup>, la rilevazione campionaria di eventi rari richiede strategie particolari, sia dal punto di vista della definizione del piano di campionamento, sia da quello delle tecniche di rilevazione. In particolare, da un lato la convergenza alla distribuzione normale delle statistiche campionarie è molto più lenta (Cochran, 1977), dall'altro la numerosità campionaria, calcolata determinando a priori un livello accettabile di errore di campionamento, tende rapidamente ad assumere valori elevatissimi.

A titolo di esempio, nel caso di un campione casuale semplice delle circa 22.000.000 di famiglie italiane, ipotizzando come accettabile che il coefficiente di variazione dello stimatore  $p$  (percentuale di acquirenti), che indicheremo con  $CV(p)$  assuma un livello pari a 0.02 (2%), la numerosità campionaria varia sensibilmente in funzione del valore assunto dalla frequenza d'acquisto  $p$ . In particolare, per una selezione di diversi  $p$  si ottengono le numerosità campionarie riportate nella tavola 3. Si osserva che la numerosità attuale del campione trimestrale, pari a circa 9000 famiglie appare adeguata per la stima della frazione di acquirenti solo per quei beni che, nell'universo, registrano una

frequenza superiore al 20%. L'indicazione è ovviamente parziale poiché lo schema di campionamento a due stadi attualmente in uso consente con ogni probabilità, un guadagno di efficienza, tuttavia, le differenze che si osservano nelle numerosità campionarie corrispondenti ai livelli più bassi di  $p$  (fino a 100 volte superiori), consentono di affermare che lo schema di campionamento attuale mal si adatta alla stima delle spese rare. Occorre, infatti ricordare che il 62% delle voci di spesa, che complessivamente rappresentano poco meno del 30% della spesa sono acquistate da meno del 20% di famiglie e che, inoltre, il 43% dei beni (pari al 14% della spesa totale) sono acquistati da meno del 5% delle famiglie.

In letteratura<sup>10</sup> si trovano diverse indicazioni circa tecniche di campionamento alternative da utilizzare per la rilevazione di eventi rari, dal semplice sovracampionamento al campionamento a rete, all'*inverse sampling*. Tali soluzioni dovrebbero essere ovviamente diversificate per le diverse spese da rilevare, e quindi occorrerebbe definire una tipologia di stimatori in grado di integrare stime campionarie provenienti da piani di campionamento diversi, dato che l'obiettivo di ottenere una stima per il totale della spesa familiare rimane prioritario. Questo modo di impostare il problema conduce evidentemente a definire un piano di campionamento complesso che è un tema che va incontrando un interesse crescente nella letteratura di teoria dei campioni<sup>11</sup>.

E' possibile individuare anche un'altra soluzione al problema, facendo perno non sull'incremento del numero di famiglie da intervistare, ma sull'ampliamento del periodo di riferimento delle spese rilevate. In parte è ciò che si fa già ora nell'indagine attuale, utilizzando un intervallo di tempo trimestrale per la registrazione degli acquisti di beni durevoli. Tuttavia, il riferimento temporale più ampio appare unicamente come un *escamotage* operativo per risolvere il problema, infatti i dati rilevati su base trimestrale non subiscono un trattamento di stima diverso da quelli rilevati su base mensile o decadale, anzi, con una operazione per molti versi arbitraria, la spesa trimestrale viene riproporzionata su base mensile tramite una semplice divisione per tre, nell'ipotesi, veramente eroica, di una sostanziale omogeneità negli acquisti che si verificano nei tre mesi. Inoltre anche ampliando a tre mesi il periodo di osservazione delle spese, la maggior parte degli acquisti rari rimane tale, il ciclo di vita dei beni durevoli è sicuramente al di sopra di questo intervallo temporale, tuttalpiù il trimestre potrebbe costituire una base sufficiente per la rilevazione degli acquisti alcuni beni immagazzinabili.

Non è chiaro quali ipotesi stiano alla base di questo modo di procedere, in particolare quali considerazioni portano alla scelta del trimestre e quali criteri determinano l'insieme dei beni il cui acquisto viene rilevato su base trimestrale. Occorre anche sottolineare che le merci incluse fra i beni durevoli sono in numero piuttosto esiguo, e di certo non comprendono tutti i beni ad acquisto raro.

Se si osservasse un intervallo temporale più ampio il problema della rarità degli acquisti troverebbe una soluzione ideale, tuttavia, utilizzando, come si fa ora, una unica intervista, è evidente che quanto più si estende l'intervallo, tanto più la rilevazione dell'acquisto corre il rischio di presentare errori dovuti alla memoria del rispondente a cui si chiede di ricostruire retrospettivamente le spese effettuate.

Appare evidente, date le considerazioni fin qui svolte, che la seconda possibile soluzione al problema delle spese rare è il ricorso ad indagini ripetute nel tempo, che consentono di seguire il processo di spesa dei nuclei familiari per periodi piuttosto lunghi. Le indagini di tipo panel, che sono già in uso in alcuni paesi per la rilevazione dei consumi familiari<sup>12</sup>, non sono tuttavia esenti da problemi che sono stati ampiamente studiati<sup>13</sup> e che non riprenderemo compiutamente in questa sede, limitandoci a considerare solo quelli che più direttamente potrebbero influire sulla efficacia della rilevazione delle spese rare.

In primo luogo la gestione di un campione *panel* richiede un notevole sforzo organizzativo e di solito consente un risparmio di numerosità campionaria; in altri termini una indagine *panel* di norma coinvolge un numero minore di unità campione rispetto a quelle condotte con campioni indipendenti ripetuti. In secondo luogo è nota la difficoltà di mantenere nel *panel* le unità campione per periodi di tempo molto lunghi.

Queste due caratteristiche considerate simultaneamente hanno una notevole influenza sul problema della rilevazione delle spese infrequenti. Infatti, non vi è dubbio che vi sono alcune spese che hanno ciclo poliennale che difficilmente potrebbe essere coperto dal ciclo di interviste del panel, tali spese beneficerebbero solo molto marginalmente della possibilità di seguire nel tempo le famiglie e, nel contempo, soffrirebbero della ridotta numerosità connessa all'uso della indagine longitudinale.

#### 4. Le frequenze di spesa e la stima di modelli di comportamento individuali

Le considerazioni svolte nei paragrafi precedenti hanno consentito di evidenziare le difficoltà del piano di campionamento attualmente in uso nel fornire misure adeguate del consumo familiare, in presenza di comportamenti di acquisto che mostrano cadenze temporali diversificate e soprattutto quando l'acquisto di un bene si configura come un evento raro. Questa difficoltà si riverbera in modo consistente sul piano delle analisi statistiche ed economiche che traggono origine dai dati così rilevati. In particolare il punto cruciale del problema è la impossibilità di attribuire un significato univoco ai valori nulli di spesa che si osservano. In forza di quanto fin qui argomentato essi possono indicare, indifferentemente sia una effettiva assenza di consumo, sia una semplice assenza di acquisto, dovuta alla relativa ristrettezza del periodo di osservazione. Si tratta, evidentemente di un limite piuttosto severo, soprattutto quando l'obiettivo dell'analisi è focalizzato sulla interpretazione dei comportamenti individuali di consumo poiché con ogni probabilità, la variabilità individuale nei comportamenti di acquisto è assai elevata.

Il problema è particolarmente sentito dagli studiosi che si sono proposti di definire la struttura e i parametri di un modello di comportamento individuale di consumo, tanto che questo argomento, che, come abbiamo già accennato, viene denominato il "problema degli zeri di spesa", compare in quasi tutti i contributi sulla stima di funzioni di comportamento a partire da dati individuali.

Gli studi in cui si è cercato di dare sistemazione teorica e soluzione empirica al problema si richiamano, da un lato al modello Tobit (Tobin, 1958; Cragg, 1971; Heckmann, 1980; Deaton e Irish, 1984), e, dall'altro, a modelli che trattano il problema in termini di probabilità di spesa (Kay, Keen e Morris, 1984; Blundell e Meghir, 1986).

##### 4.1 La probabilità di spesa nel modello Tobit

Per affrontare il problema della frequenza di spesa Tobin propone un modello che coniughi le proprietà di un'analisi probit con quelle di un modello di regressione multipla. L'obiettivo è definire un modello uni/multiequazionale in cui la variabile dipendente, cioè il consumo del bene osservato per una data famiglia, risulti correlata, oltre che alle caratteristiche proprie della famiglia, anche alla probabilità di osservare bassi livelli di spesa e alla numerosità delle

spese che differiscono da tali estremi inferiori. Secondo le ipotesi formulate da Tobin, i livelli più bassi di spesa starebbero ad indicare, per ogni bene o categoria di beni, quegli acquisti non danno ragione del consumo effettuato.

Per i beni rari, tale limite inferiore è ovviamente pari a zero poiché, nel limitato periodo in cui vengono registrate le spese, poche famiglie acquisteranno quei beni, pur potendo continuare a consumare beni acquistati in precedenza. Tuttavia il limite più basso di spesa non è necessariamente zero; infatti anche spese non nulle per un bene possono misurare in maniera inadeguata il suo consumo. Tobin definisce, almeno nella presentazione del modello, tale limite inferiore diverso da zero, correlandolo con il livello del reddito familiare. Anche se intuitivamente accettabile la definizione di questo limite pare, nella forma proposta da Tobin, difficilmente identificabile; tanto che, le specificazioni di tale modello proposte negli ultimi anni si limitano a considerare come limite inferiore di spesa lo zero.

Fra i contributi appena richiamati particolare attenzione va posta al lavoro di Deaton e Irish che propongono un'integrazione del modello di Tobin introducendo un processo censorio che serve a rimpiazzare casualmente una frazione delle osservazioni generate dal modello Tobit con degli zeri. Il loro obiettivo è poter distinguere gli effettivi mancati acquisti da una variegata tipologia di errate registrazioni. In altri termini, oltre agli zeri di spesa legati alla bassa frequenza di acquisto, vengono prese in considerazione anche altre concause quali: il rifornimento, nel periodo di indagine, di forti quantità di bene che saranno consumate anche in tempi successivi; errori di compilazione dei diari di spese; il fenomeno che Kemsley e altri (1980) denominano *beginning of period effects* che consiste nella possibilità che alcuni consumatori che hanno effettuato acquisti appena prima del periodo di osservazione, vogliano, per così dire, aiutare l'indagine registrando ugualmente l'acquisto e quindi spostandolo nel tempo.

Il modello di Deaton e Irish è stato applicato ai dati dell'indagine sui consumi delle famiglie inglesi del 1974, per la stima delle determinanti del consumo di tabacco, alcool e alcuni beni durevoli. Come variabili indipendenti sono stati assunti indicatori economici (spesa familiare pro-capite) e demografici (composizione della famiglia ed età dei suoi membri), condizione lavorativa, possesso di alcuni beni durevoli (auto, telefono, abitazione) e regione di residenza.

Se il consumo non osservabile è sempre positivo, il modello fornisce una specificazione perfettamente adeguata e può agevolare la definizione

determinanti del comportamento di consumo per alcuni beni, esso ha, inoltre, il vantaggio di essere compatibile con una distribuzione della componente stocastica priva di punti di discontinuità, ad esempio normale.

Per valutare i risultati dell'applicazione del modello di Deaton e Irish sono stati utilizzati dei test derivati dai moltiplicatori di Lagrange e sviluppati, tra gli altri, da Blundell e Megh   (1986) e Gouri  roux, Monfort, Renault e Trognon (1984). Tali test hanno portato a rifiutare tutte le conclusioni a cui si era giunti con l'applicazione del modello Tobit. Una causa di questi risultati poteva essere l'assunzione di normalit   del processo di sostituzione casuale dei dati mancanti, propria del modello Tobit; Deaton e Irish hanno allora sondato la possibilit   di ristimare il modello Tobit senza la condizione di normalit   ricorrendo alle procedure di stima non parametrica di Buckley e James (1979), ma i risultati, sorprendentemente, non sono stati molto diversi

#### 4.2 Il comportamento di spesa come processo stocastico

A differenza dei lavori appena proposti i contributi di Kay, Keen e Morris e di Blundell e Meghir delimitano decisamente la loro attenzione al problema degli zeri di spesa legati alla bassa frequenza di acquisto. In questo contesto, esplicita attenzione viene posta al problema delle determinanti della probabilit   di acquisto di un bene.

In altri termini, per Kay, Keen e Morris, la definizione di un modello di comportamento di consumo implica l'approfondimento della dell'analisi della struttura stocastica del processo di spesa. Tale stocasticit   emerge con grande evidenza anche da una semplice ispezione dei dati, che sono caratterizzati da un'alta incidenza di zeri ad ogni livello di aggregazione dei beni. Questa peculiare presenza di zeri non pu   essere attribuita, secondo gli autori, interamente a variazioni nelle preferenze dei consumatori e/o a limiti del processo di rilevazione. Deve quindi essere ipotizzata una sorta di indeterminatezza stocastica nel processo di consumo, che in questo senso sarebbe definito in uno spazio parametrico continuo, mentre la spesa effettivamente osservata ne costituirebbe una determinazione discreta. In questo senso il consumo pu   essere definito come una variabile latente di cui la spesa    manifestazione parziale e affetta da errori

E' soprattutto nel lavoro di Blundell e Meghir che questo approccio viene sviluppato e che la probabilit   di acquisto viene inserita esplicitamente nel modello. Pi   precisamente l'obiettivo strumentale    spiegare la distribuzione

degli acquisti collegando la probabilit   d'acquisto al reddito e ad altre caratteristiche familiari in modo da minimizzare il peso delle variabili latenti (ad esempio errori di registrazione o *beginning of period effect*). L'obiettivo ultimo    la stima della curva di Engel che rappresenta la relazione che lega il consumo al reddito ed ad altre caratteristiche del nucleo familiare. Va detto che l'autore mostra scarsa attenzione al problema dell'identificazione un sistema completo di domanda, e si limita all'adattamento del suo modello solo ad alcuni beni. In particolare una applicazione del modello    stata condotta sulla spesa per abbigliamento. La selezione delle variabili indipendenti, con cui descrivere la curva di Engel e soprattutto definire la probabilit   d'acquisto,    stata compiuta utilizzando molte delle indicazioni provenienti dai lavori precedentemente presentati dagli stessi Autori e da altri quali Blundell (1980), Blundell e Walker (1982) e Blundell e Ray (1984). Da questi studi apparivano fra le determinanti del consumo variabili quali la numerosit   e la composizione della famiglia, il reddito e il salario. In aggiunta a queste, gli Autori hanno ritenuto opportuno utilizzare come variabili il grado d'istruzione e la condizione lavorativa.

I risultati paiono, almeno questa volta, accettabili. Soddisfacenti risultano essere le propriet   delle stime dei parametri sia della curva di Engel che delle frequenze d'acquisto. Anche le assunzioni circa la normalit   e indipendenza delle probabilit   di acquisto, verificate usando gli appropriati test dei moltiplicatori di Lafrange, si mostrano accettabili. Pi   in generale l'impatto complessivo del reddito e delle variabili economiche e demografiche    stato trovato significativo e plausibile.

#### 5. Le frequenze di spesa e la ricerca di tipologie di consumatori

L'influenza delle basse frequenze di spesa sulla individuazione di tipologie di comportamento omogeneo nel consumo delle famiglie trova pochissimi riferimenti nella letteratura statistico-economica. Per di pi  , nei pochi riferimenti disponibili che possono essere ritrovati in alcuni lavori della scuola dell' *Analyse des donnees* (Lebart, Morineau e Tabard, 1978; Glaude e Mourtadier, 1982), il problema emerge solo marginalmente e viene trattato come inconveniente tecnico da risolvere con soluzioni operative. Cos   nel lavoro di Lebart, Morineau e Tabard, che ha come obiettivo l'individuazione di gruppi omogenei di famiglie consumatrici, le uniche spese prese in esame sono

quelle per i beni alimentari, correnti e semidurevoli che, nel caso citato, rappresentavano solo circa il 50% delle spese totali. Le spese per i beni durevoli non sono state tenute in considerazione proprio perché si riteneva inadeguato il piano dell'indagine per cogliere le spese rare<sup>14</sup>. Tuttavia, proprio queste spese, più di quelle frequenti, hanno potere discriminante nella definizione di tipologie omogenee di consumatori, essendo più consistenti e più diversificate in funzione dello stile di vita e dell'abitudine di consumo.

Le informazioni sui beni durevoli sono, in un secondo tempo, utilizzate come aiuto alla interpretazione del comportamento di consumo dei gruppi omogenei di famiglie; così facendo, però, tali spese vengono trattate alla stregua di una qualsiasi altra variabile strutturale caratterizzante la famiglia. Le spese per durevoli, dunque non contribuiscono alla individuazione dei gruppi omogenei, e vengono utilizzate solo in chiave descrittiva.

D'altra parte molto scarsa è l'attenzione che viene prestata in letteratura al trattamento formale della presenza di spese rare e in particolare agli effetti che la loro presenza ha sugli algoritmi di calcolo che di norma sono utilizzati per l'analisi multivariata su cui si impernia la ricerca di tipologie omogenee. Nel seguito del paragrafo si affronta un aspetto peculiare di questo problema, in particolare si è studiato verrà evidenziato l'effetto della presenza di frequenze di acquisto diverse su due strumenti statistici cardine delle analisi multivariate: la matrice di varianza-covarianza e la matrice di distanza.

L'elaborazione della matrice di varianza e covarianza relativa all'ammontare delle spese per i diversi beni è il nucleo di tutti gli algoritmi che fanno capo all'analisi fattoriale finalizzata alla ricerca di tipologie di consumo. In questo contesto esplicitare il ruolo delle frequenze d'acquisto nella misura della covarianza consente di mostrare come la variabilità e la covariabilità che emerge dalla osservazione dell'ammontare complessivo della spesa per due beni qualsiasi, è frutto di diverse componenti distinte: la diversità nella proporzione di acquirenti in ciascuno dei due beni, la variabilità della spesa nel dominio degli acquirenti di ciascun bene, la covarianza delle spese per i due beni nel dominio degli acquirenti simultanei delle due merci.

Più in dettaglio, dato un collettivo di riferimento di  $n$  famiglie, e le spese

per l'acquisto di due beni (a e b), siano:

$n_a$  = numero di famiglie acquirenti del bene a

$n_b$  = numero di famiglie acquirenti del bene b

$n_{na}$  = numero di famiglie non acquirenti del bene a

$n_{nb}$  = numero di famiglie non acquirenti del bene b

$y_{ia}$  = ammontare della spesa per il bene a, nella famiglia i

$y_{ib}$  = ammontare della spesa per il bene b, nella famiglia i

$\bar{y}_a$  = spesa media per il bene a nel dominio degli acquirenti del bene a

$\bar{y}_b$  = spesa media per il bene b nel dominio degli acquirenti del bene b

$n_{a,b}$  = numero di famiglie che acquistano simultaneamente il bene a e il bene b

$n_{na,nb}$  = numero di famiglie che non acquistano simultaneamente il bene a e il bene b

$n_{a,nb}$  = numero di famiglie che acquistano il bene a e non il bene b

$n_{na,b}$  = numero di famiglie che acquistano il bene b e non il bene a

La covarianza fra le spese relative ai beni a e b è:

$$Cov_{ab} = \frac{\sum_{i=1}^n (y_{ia} - \bar{y}_a)(y_{ib} - \bar{y}_b)}{n}$$

Il numeratore può essere facilmente scomposto in quattro componenti:

a) codevianza calcolata sulle famiglie acquirenti di ambedue i beni

$$A = \sum_{i=1}^{n_{a,b}} (y_{ia} - \bar{y}_a)(y_{ib} - \bar{y}_b)$$

b) codevianza calcolata sulle famiglie non acquirenti di entrambi i beni

$$B = n_{na,nb} \bar{y}_a \bar{y}_b$$

c) codevarianza calcolata sulle famiglie acquirenti il bene a e non acquirenti il bene b

$$C = n_{a,nb} \bar{y}_b \sum_{i=1}^{n_{a,nb}} (y_{ia} - \bar{y}_a)$$

d) codevarianza calcolata sulle famiglie non acquirenti il bene a e acquirenti il bene b

$$D = n_{na,b} \bar{y}_a \sum_{i=1}^{n_{na,b}} (y_{ib} - \bar{y}_b)$$

vale, ovviamente, la relazione:

$$Cov_{ab} = \frac{Codev_{ab}}{n} = \frac{A + B + C + D}{n_{a,b} + n_{na,nb} + n_{a,nb} + n_{na,b}}$$

L'espressione precedente consente di isolare i contributi alla covarianza delle diverse tipologie di acquirenti e non acquirenti, infatti, il primo termine esprime la variabilità congiunta delle spese nel dominio degli acquirenti di entrambi i beni in esame, il secondo rappresenta l'effetto dei singoli valori medi di spesa calcolati nel dominio delle famiglie che non acquistano nessuno dei due beni, infine il terzo e il quarto esprimono il contributo del sottoinsieme delle famiglie che acquistano un bene, ma non l'altro.

In sintesi, si rivelano influenti tutti i sottoinsiemi definiti dalle combinazioni di tutti i possibili comportamenti di spesa. Come è logico attendersi lo sviluppo dell'indicatore di covarianza segue la logica dello sviluppo del quadrato di un binomio e quindi i comportamenti si combinano secondo una legge moltiplicativa. Ciò ha una implicazione importante sulla forma della varianza e covarianza delle spese per i diversi beni, infatti gli effetti di un bene ad acquisto raro si ripercuotono su un gran numero di elementi della matrice di varianza e covarianza. In una analisi multivariata il problema della rarità dell'acquisto riguarda tutte le combinazioni di variabili che contengono un bene a bassa frequenza.

Per evidenziare l'effetto della rarità sul calcolo della covarianza, sono stati prescelti 6 beni che rappresentano diversi livelli di frequenza e, per ogni coppia, sono state calcolate distintamente le quattro componenti di covarianza citate in precedenza. Di seguito è riportato l'elenco dei beni prescelti

unitamente alla percentuale di famiglie che ne hanno registrato l'acquisto nel IV trimestre 1989.

a)spese ad alta frequenza:	PANE	98.7%
	MELE E PERE	88.4%
b)spese a media frequenza:	CALZATURE	48.5%
	BARBIERE PARRUCCHIERE	49.6%
c. spese a bassa frequenza:	AUTO NUOVA	1.7%
	RADIO GRAMMOFONI	1.2%

Una sintesi dei risultati del calcolo sono riportati nel prospetto che segue nel quale sono indicate le componenti (identificate con le lettere maiuscole utilizzate in precedenza) che hanno peso predominante per ciascuna coppia di tipologie di beni.

**Prospetto 1 - Componente di codevarianza dominante per tipologia di beni**  
(derivate dai valori assoluti di codevarianza)

Beni	ad alta frequenza	a media frequenza	a bassa frequenza
ad alta frequenza	A	A	A
a media frequenza	A	A - C/D	A - D/C
a bassa frequenza	A	A - D/C	A

Come è ovvio, in generale, ciò che emerge è che la componente di gran lunga più importante è costituita dalla covarianza degli acquirenti(A) simultanei di ciascuna coppia di beni quando i beni appartengono alla stessa categoria di frequenza (rari o frequenti). Tuttavia la valutazione è resa

problematica dal fatto che i diversi sottoinsiemi di famiglie considerati hanno numerosità molto diversa e il peso relativo di ciascuna componente risente di questa diversa numerosità. Ad esempio se confrontiamo pane e mele il peso preponderante della componente A è dovuta al fatto che un gran numero di famiglie compra sia pane che mele. E' utile, quindi standardizzare la misura del contributo alla codevarianza complessiva rapportando il valore delle singole componenti al numero di famiglie del sottoinsieme di competenza. Questa sorta di contributo medio alla codevarianza complessiva mostra un comportamento molto particolare e consente di evidenziare con chiarezza l'influenza delle diverse frequenze di acquisto.

**Prospetto 2 - Componente di codevarianza dominante per tipologia di beni**  
(derivate dai valori standardizzati di codevarianza)

Beni	ad alta frequenza	a media frequenza	a bassa frequenza
ad alta frequenza	B	B-C	D
a media frequenza	B-C	A-B-C-D	A-D
a bassa frequenza	D	A-D	A

Ai due estremi del prospetto 2 si collocano le componenti che identificano gli acquirenti e non acquirenti simultanei dei due beni, le famiglie che non acquistano beni frequenti hanno mediamente un peso determinante nella misura della relazione fra spese frequenti mentre, al contrario, per gli acquisti rari, è la simultaneità dell'acquisto che presenta l'importanza maggiore. Nella misura della relazione fra beni a media frequenza le quattro componenti giocano un ruolo paritario. Nelle caselle al di fuori della diagonale principale, si osserva che è sempre il sottodominio di famiglie di numerosità minore che emerge: scorrendo le colonne, si osserva che per le spese frequenti, sono i non

acquirenti del bene frequente che pesano proporzionalmente di più, mentre per i beni rari, sono le famiglie acquirenti a dominare.

Considerazioni del tutto parallele si evidenziano esaminando la definizione della misura principale su cui si fonda l'analisi di raggruppamento: la distanza fra due individui. Se si considera, ad esempio, come indicatore il quadrato della distanza euclidea, date due famiglie (i e j) ed m beni, siano:

$$y_i = \begin{cases} y_{ia} & i \in A \\ 0 & i \in B \end{cases}$$

$y_{ia}$  = spesa della famiglia i per l'acquisto del bene a

Il quadrato della distanza euclidea fra le famiglie i e j è:

$$D_{ij}^2 = \sum_{a=1}^m (y_{ia} - y_{ja})^2 =$$

$$= \sum_{a=1}^m f_{ia}^2 y_{ia}^2 + \sum_{a=1}^m f_{ja}^2 y_{ja}^2 - 2 \sum_{a=1}^m f_{ia} y_{ia} \sum_{a=1}^m f_{ja} y_{ja}$$

Anche in questo caso è evidente l'importanza della presenza di spese congiunte. Ovviamente il punto di vista è simmetrico rispetto al precedente: l'insieme-intersezione da considerare è costituito dalle voci di spesa presenti simultaneamente nel profilo di acquisto delle due famiglie.

Infatti, se si considerano i tre addendi in cui è stata scomposta la distanza fra le due famiglie, appare evidente che i primi due non riguardano l'accoppiamento tra le famiglie, sono cioè legati unicamente al comportamento singolo di ciascun nucleo familiare, mentre il prodotto delle spese congiunte (il terzo addendo) è l'unico che dipende simultaneamente dai profili di spesa delle due unità considerate, ed è quindi quello che, in ultima analisi determina la distanza. E' quindi il numero e la consistenza delle spese comuni che determina la somiglianza o la diversità tra i comportamenti di spesa delle famiglie.



## 6. Considerazioni conclusive

Siamo ben consci che le argomentazioni presentate in queste pagine mal si prestano a formulare considerazioni di carattere conclusivo. I diversi aspetti che sono stati trattati, in realtà, aprono questioni e pongono interrogativi, più che fornire risposte. D'altra parte, l'obiettivo principale di questa fase del lavoro di analisi sui dati dell'Indagine BF è costituito proprio dallo stimolo ad una riflessione critica sull'accuratezza, coerenza e qualità dei dati raccolti. Su questo piano alcune evidenze sono emerse con chiarezza.

In primo luogo, occorre tenere presente la relazione complessa e mutevole che lega spesa e consumo e le conseguenze che questa relazione ha sul processo di misura, in particolare sulla individuazione di coerenti strategie di rilevazione degli acquisti sia dal punto di vista del piano di campionamento, sia per quanto concerne le procedure di raccolta dei dati.

In secondo luogo, è risultato evidente che la qualità dei dati o, più precisamente, il significato attribuibile all'assenza di spesa ha forte impatto sulle tecniche econometriche e statistiche che vengono utilizzate per l'analisi; in particolare, gran parte delle usuali assunzioni su cui si reggono tali metodologie appaiono violate e, quindi, vi è la possibilità di conseguire risultati inconsistenti.

I passi successivi che l'analisi richiede dovranno, coerentemente, essere indirizzati in una duplice direzione. Sul piano dell'indagine, verso la definizione di strategie di campionamento e di raccolta dei dati in grado di rilevare le informazioni sugli acquisti rari, riducendo al minimo le distorsioni e gli inconvenienti legati all'ampliamento del periodo di osservazione e alla ripetizione dell'indagine sugli stessi soggetti. Sul piano dell'analisi econometrica dei dati individuali, appare irrinunciabile il ricorso a tecniche di stima in grado di trattare il problema delle spese nulle. Infine, un ampio e pressoché inesplorato territorio di ricerca si apre per quanto concerne l'analisi delle tipologie di consumo attraverso tecniche statistiche multivariate. La definizione di metriche adeguate e di algoritmi adatti a trattare il problema degli zeri di spesa è piuttosto trascurato in letteratura. Su questo terreno è nostra intenzione tentare di percorrere un piccolo tratto di cammino.

## Note

- 1 Una descrizione approfondita dell'Indagine è riportata in Innocenzi (1992)
- 2 Il tema è affrontato in questa stessa collana da Filippucci, Marliani (1992)
- 3 Le definizioni riportate sono piuttosto generali e comuni a tutti i manuali di economia, si può citare, ad esempio, Zamagni (1979) e Deaton, Muellbauer (1980)
- 4 Si veda, ad esempio, Deaton, Muellbauer (1980) e Lancaster (1966)
- 5 Si veda ad esempio, Blundell, Meghir (1986) e Pudney (1990)
- 6 L'analisi delle frequenze di spesa per tipologia familiare è riportata in Drudi, Filippucci, Messori (1992)
- 7 Zeri di spesa è la traduzione dell'espressione "*zero expenditures*" comune a quasi tutti gli autori citati nel testo
- 8 Per analisi comparativa dei dati delle fonti BF e CN si veda Mantegazza, Tassinari (1991)
- 9 Si veda Cochran (1977) e Cicchitelli, Herzel, Montanari (1992)
- 10 Una sintetica descrizione della tecnica dell'*inverse sampling* e della bibliografia relativa può essere ritrovata in Cochran (1977)
- 11 Per una rassegna della problematica dei campioni complessi si veda Chaudhuri, Vos (1988) e Skinner, Holt, Smith (1989)
- 12 Per una proposta di indagine panel in tema di rilevazione dei consumi in Italia si veda Lalla (1992)
- 13 Un'esauriente rassegna di tali problemi è riportata in Duncan, Kalton, Kasprzyk, Singh (1989)
- 14 L'analisi si riferisce ai dati raccolti nel 1963 e a quel tempo la metodologia dell'indagine INSEE era molto diversa da quella attuale.

## Riferimenti Bibliografici

- BLUNDELL R.W., *Estimating continuous consumer equivalent scales in an expenditure model with labour supply*, in "European Economic Review", 40, pp. 145-157, 1980
- BLUNDELL R.W., MEGHIR C., *Bivariate alternatives to the tobit model*, in "Journal of Econometrics", 1986
- BLUNDELL R.W., MEGHIR C., *Selection criteria for a microeconomic model of labour supply*, in "Journal of Applied Econometrics", 1, pp. 57-84, 1986
- BLUNDELL R.W., MEGHIR C., *Engel curve estimation with individual data*, in Heijmans R. -Neudecker H. (eds.), *The practice of econometrics: Studies on demand, forecasting, money and income*, Kluwer Academic Publishers, Dordrecht, 1987
- BLUNDELL R.W., RAY R., *Testing for linear Engel curves in a new flexible demand system*, in "Economic Journal", Conference Papers, 1984
- BLUNDELL R.W., WALKER I., *Modelling the joint determination of household labour supplies and commodity demands* in "Economic Journal", 42, pp. 351-364, 1982
- BROWN J.A.C., DEATON A.S., *Models of consumer Behaviour: A Survey*, in "Economic Journal", 82, pp. 1145-1236, 1972
- BUCKLEY J., JAMES I., *Linear regression with censored data*, in "Biometrika", 66, pp. 429-436, 1979
- CHAUDURI A., VOS J.W.E., *Unified theory and strategies of survey sampling*, North Holland, Amsterdam, 1988
- CICCHITELLI G., HERZEL A., MONTANARI G.E., *Il campionamento statistico*, Il mulino, Bologna, 1992
- COCHRAN W.G., *Sampling techniques*, III ed., J. Wiley, New York, 1977
- CRAGG J.G., *Some statistical models for limited dependent variables with applications to the demand for durable goods*, in "Econometrica", 39, pp. 829-844, 1971
- DEATON A.S., IRISH M., *A statistical model for zero expenditures in household budgets*, in "Journal of public economics", 23, 1984
- DEATON A., MUELLBAUER J., *Economics and consumer behaviour*, : Cambridge University Press, Cambridge, 1980
- DRUDI I., FILIPPUCCI C., MESSORA E., *Tipologie di consumo e determinanti del consumo individuale*, CON.PRI. Rapporto di ricerca n.21, Dipartimento di Scienze Statistiche, Bologna, in corso di pubblicazione

- FILIPPUCCI C., MARLIANI G., *La misura dei consumi delle famiglie: una riflessione a partire dall'esperienza italiana*, CON.PRI. Rapporto di ricerca n.6, Dipartimento di Scienze Statistiche, Bologna, 1992
- GLAUDE M., MOURTARDIER M., *Les budgets des menages*, in "Economie et statistique", 1982
- GOURIEROUX C., MONFORT A., REANAULT E., TROGNON A., *Residus generalists ou interpretations lineaires de l'econometrie non lineaire*, INSEE, document de travail, 8410, 1984
- GREENE W.H., *On the asymptotic bias of the ordinary least squares estimation of the tobit model*, in "Econometrica", 49, pp. 505-513, 1981
- HECKMAN J.J., *Sample selection bias as a specification error*, in Smith J.P. ed., *Family labor supply*, Princeton University Press, 1979
- HECKMAN J.J., MACURDY T., *A life-cycle model of female labor supply*, in "Review of economic Studies", pp. 47-74, 1980
- INNOCENZI G., *Aspetti della rilevazione e dell'elaborazione dei dati dell'indagine sui consumi delle famiglie*, CON.PRI. Rapporto di ricerca n.2, Dipartimento di Scienze Statistiche, Bologna, 1992
- ISTAT, *I consumi delle famiglie - anno 1988*, in "Collana d'informazione", 30, Roma, 1990
- KAY J.A., KEEN M.J., MORRIS C.N., *Consumption, income and interpretation of household expenditure data*, in "Journal of public economics", 23, pp. 69-181, 1984
- DUNCAN G., KALTON G., KASPRZYK D., SINGH M.P., *Panel surveys*, Wiley, New York, 1989
- KEMSLEY W.F.F., REDPAT R.U., HOLMES M., *Family expenditure survey handbook*, HMSO, London, 1980
- LALLA M., *Proposte alternative alla rotazione completa nell'indagine sui consumi delle famiglie*, CON.PRI. Rapporto di ricerca n.5, Dipartimento di Scienze Statistiche, Bologna, 1992
- LANCASTER K., *A new approach to consumer theory*, in "Journal of Political Economy", vol. LXXIV, 1966
- LEBART L., MORINEAU A., TABARD N., *Tecniche de la description statistique*, 1978
- MANTEGAZZA S., TASSINARI G., *La coerenza delle valutazioni dei consumi privati nella contabilità nazionale e nell'indagine sui bilanci delle famiglie*, CON.PRI. Rapporto di ricerca n.4, Dipartimento di Scienze Statistiche, Bologna, 1992

- MORINEAU A., *Note sur la caracterisation statistique d'une classe et les valeurs-tests*, in "Bulletin Technique du CESIA", Vol 2, Paris, 1984
- MUSGROVE P., *Consumer behaviour in Latin America: income and spending of families in ten andean cities.*, Brookings , Washington, 1978
- PUDNEY S.E., *The estimation of Engel curves*, in MYLES G.D. (ed.), *Measurement and modelling in economics*, North Holland, Amsterdam, 1990
- SKINNER C.J., HOLT D., SMITH T.F.F., *Analysis of complex surveys*, Wiley, Chichester, 1989
- THEIL H., *Theory and measurement of consumer demand*, Vol I - Vol II, North-Holland, 1976
- TOBIN J., *Estimation of relationships for limited dependent variables*, in "Econometrica" , 26, pp. 24-36, 1958
- ZAMAGNI S., *Georgescu Roegen. I fondamenti della teoria del consumatore*, Etas libri, Milano, 1979

## Appendice

Tavola 1 - Percentuale di famiglie acquirenti per categoria di spesa

Categorie di spesa	Percentuale acquirenti	Categorie di spesa	Percentuale acquirenti
PANE	98.73	GAS DA RETE	52.48
ENEL	98.65	MAIALE	50.17
LATTE	92.81	MARMELLATA CACAO	49.69
FORMAGGI	90.98	BARBIERE PARUCCHIERE	49.58
PASTA	89.72	OLIO SEMI	48.61
MELE PERE	88.36	CALZATURE	48.52
SALUMI	85.52	MEDICINALI	47.49
LEGUMI ORTAGGI FRESC	85.21	POMODORI FRESCI	47.43
TELEFONO	83.67	SPESE CONDOMINIALI	46.18
UOVA	81.94	TABACCHI	45.17
BISCOTTI PASTICCERIA	81.25	POMODORI IN SCATOLA	43.44
CAFFE TE	79.76	LEGUMI ORTAGGI CONS.	41.00
SAPONE DETERSIVI	79.12	SCOPE FIAMMIFERI	39.07
SAPONETTE DENTIFRICI	77.89	DENARO AI FIGLI	36.65
AGRUMI	77.52	RISTORANTI PIZZERIE	35.28
VITELLO	75.63	LAVANDERIA TINTORIA	35.16
FITTO FIGURATIVO	74.13	PESCE CONSERVATO	34.01
GIORNALI RIVISTE	73.30	ACQUA	32.68
ZUCCHERO	72.16	PIANTE E FIORI	30.41
POLLAME	72.05	TOTOCAL LOTTO ALTRI	30.37
BENZINA AUTO MOTO	69.96	BIGL. ABB. AUTOBUS	29.68
PESCE FRESCO	68.23	ABITI CAPPOTTI	29.63
PATATE	67.84	CINEMA TEATRO DISCOT	27.53
OLIO OLIVA	64.47	GAS BOMBOLE	26.66
RISO	61.84	FRANC. SPESE POSTALI	26.05
BAR	61.29	AFFITTO	25.84
CALZE MERCERIA	61.18	CALZOLAIO	25.39
VINO	58.73	QUADERNI CANCELLERIA	25.13
SALE ACETO ETC.	57.78	ABB.SPORTIVO E T.L.	24.93
MANZO	57.41	FARINA	24.83
BURRO	57.37	BIANCHERIA CASA	23.60
ACQUA MINERALE	54.96	MAGLIERIA E TESSUTI	22.38

segue Tavola 1 - Percentuale di famiglie acquirenti per categoria di spesa

Categorie di spesa	Percentuale acquirenti	Categorie di spesa	Percentuale acquirenti
CAMICIE	22.20	ARGENT., GIOIELLI	6.25
MAGLIERIA INTIMA	20.57	MANUT.ELETTRODOM.	6.20
INSETT.DISINF.VERNIC	20.40	ABBON. RADIO TV	6.12
LIQUORI	19.93	COPERTE TESS.ARRED.	6.07
GIOCATTOLE	18.95	PENTOLE PADELLE POS.	5.99
BIRRA	18.16	FITTO FIGUR. II CASA	5.60
CARROZZERIA	17.86	KEROSENE GASOLIO	5.56
LIBRI	17.14	ISTRUZIONE	5.37
SPESE ANIM. DOMEST.	16.97	OVINA E CAPRINA	5.12
MARGARINA	16.94	BORSE ARTICOLI PELLE	4.93
CONGLI	15.61	BIGLIETTI ABB. FERR.	4.81
BEVANDE ANALCOLICHE	14.99	ENEL II CASA	4.66
DISCHI FONOCASSETTE	14.82	PIATTI E TAZZE	4.63
MEDICI INFERMIERI	14.18	CARNE IN SCATOLA	4.61
RISC.CENTRALE	13.93	PICCOLI APP.ELETTTR.	4.30
ASS.AUTO E ALT. VEIC	13.34	ALTRE SPESE FINANZ.	3.85
GASOLIO ALTRI COMB.	12.68	TESSUTI x CONF.MISU.	3.79
SVILUPPO PELLICOLE	11.36	BICCHIERI BOTTIGLIE	3.78
ALTRE SPESE AUTO MOT	10.81	PNEUMATICI	3.77
FRUTTA SECCA	9.85	ARTICOLI PERSONALI	3.60
LAMPADARI QUADRI	8.86	FALEGNAME	3.51
MENSE	8.59	TASSE SCOLASTICHE	3.32
SUCCHI DI FRUTTA	8.43	ALTRI BENI CASA	3.23
OLII E ALTRI LUBRIF.	8.31	EQUINA	2.81
CULTO	8.19	MOBILI LEGNO	2.80
ASSICUR. VITA	7.70	ALTRE ASSICURAZIONI	2.75
CONF.Z. ARTICOLI ABB.	7.37	SERVIZI DOMESTICI	2.64
ANALISI MEDICHE	7.10	BRICOLAGE HOBBY	2.59
OCCHIALI SIRINGHE	6.93	FRUTTA CONSERVATA	2.51
CARBONE LEGNA	6.55	GARAGE AUTO MOTO	2.40
STOVIGLIE PLASTICA	6.47	TAPPETI MOQUETTE	2.01
IMPIANTI	6.25	MANUT. VESTIARIO	1.95

segue Tavola 1 - Percentuale di famiglie acquirenti per categoria di spesa

Categorie di spesa	Percentuale acquirenti	Categorie di spesa	Percentuale acquirenti
ARTICOLI SPOR. CAMP.	1.75	STRUM. MUSICALI	0.78
ACQUISTO AUTO NUOVA	1.65	LUCIDATRICI ASPIR.	0.77
MATERASSI	1.64	ANIMALI DOMESTICI	0.66
MAN. MACC. FOT. E AL.	1.62	GAS RETE II CASA	0.65
ALTRE SPESE RICREAT.	1.62	PERSONAL COMPUTER	0.60
LAVATRICI	1.49	MOBILI METALLO	0.51
BICICLETTE	1.47	ACQ. SCOOTERS MOTO	0.48
TELEVISORI COLORI	1.46	AFFITTO II CASA	0.44
ACQUA II CASA	1.44	BIGL. MEZZI MARITT.	0.41
SP.CONDOM. II CASA	1.37	RICOVERI OSPEDALIERI	0.40
ONORARI LIB. PROFES.	1.32	LAVASTOVIGLIE	0.34
MACCH.FOT.,CINEPRESE	1.29	ALTRE SPESE TRASPORT	0.33
GITE VIAGGI TURIST.	1.27	RISC.CENTRALE II CASA	0.32
RADIO RADIOGRAMMOFO.	1.22	MOBILI PLASTICA	0.31
VIDEO REGISTRATORI	1.21	CARBONE LEGNA II CASA	0.26
ALTRE MEDICHE	1.21	MACCH. DA SCRIVERE	0.24
ACQUISTO AUTO USATA	1.18	KEROS.GASOLIO II CASA	0.21
CUCINE STUFE SCALD.	1.18	BIGLIETTI AEREI	0.20
GAS BOMBOLE II CASA	1.11	IMPIANTI II CASA	0.16
ALBERGHI PENSIONI	1.02	CONDIZIONATORI	0.15
PELLICCE E PELLE	0.97	MACCH. CUCIRE	0.13
MANUT.BIANCHERIA	0.91	TELEVISORI BIANCONER	0.13
LARDO	0.91	CASE RIPOSO	0.06
FRIGORIFERI CONG.	0.85	ACQUISTO ROULOTTE	0.05
GIRADISHI REGISTRAT.	0.79	BARCHE CANOTTI	0.04

Tavola 2 - Coefficiente di variazione della stima della spesa totale e percentuale di varianza spiegata dalla rarità dell'acquisto

Categorie di spesa	Coefficiente di variazione (x100)	Percentuale varianza dovuta alla rarità
PANE	1.38	1.03
FITTO FIGURATIVO	1.52	9.94
LATTE	1.58	2.78
UOVA	1.59	5.66
MELE PERE	1.66	4.88
FORMAGGI	1.70	3.77
LEGUMI ORTAGGI FRESC	1.80	6.30
CAFFE TE	1.82	5.76
TELEFONO	1.86	4.49
ENEL	1.86	1.22
VITELLO	1.87	6.51
BENZINA AUTO MOTO	1.93	8.40
PASTA	1.94	2.78
ZUCCHERO	1.96	8.12
PESCE FRESCO	2.00	6.96
RISO	2.00	10.72
POLLAME	2.05	6.97
AGRUMI	2.07	7.94
SAPONE DETERSIVI	2.08	7.07
SALUMI	2.09	6.47
PATATE	2.10	9.68
TABACCHI	2.14	10.45
ACQUA MINERALE	2.19	9.34
BISCOTTI PASTICCERIA	2.21	7.60
BARBIERE PARUCCHIERE	2.22	13.82
CALZE MERCERIA	2.22	12.76
BURRO	2.22	13.19
BAR	2.25	9.11
MANZO	2.26	9.46
CALZATURE	2.26	10.24
POMODORI FRESCHI	2.37	14.55

segue Tavola 2 - Coefficiente di variazione della stima della spesa totale  
e percentuale di varianza spiegata dalla rarità dell'acquisto

Categorie di spesa	Coefficiente di variazione (x100)	Percentuale varianza dovuta alla rarità
OLIO SEMI	2.56	11.48
SAPONETTE DENTIFRICI	2.58	9.18
AFFITTO	2.64	22.95
GAS DA RETE	2.71	12.47
GIORNALI RIVISTE	2.71	6.46
MAIALE	2.74	9.93
MARMELLATA CACAO	2.75	10.19
LEGUMI ORTAGGI CONS.	2.77	14.27
VINO	2.81	10.75
SALE ACETO ETC.	2.82	10.91
MEDICINALI	2.84	10.44
POMODORI IN SCATOLA	2.98	12.83
OLIO OLIVA	3.08	7.62
LAVANDERIA TINTORIA	3.10	17.85
SPESE CONDOMINIALI	3.12	16.78
RISTORANTI PIZZERIE	3.13	14.02
PESCE CONSERVATO	3.16	14.33
ACQUA	3.25	23.81
CALZOLAIO	3.28	25.13
ABITI CAPPOTTI	3.37	14.73
CAMICIE	3.50	17.60
MARGARINA	3.58	25.67
MAGLIERIA E TESSUTI	3.68	19.79
MAGLIERIA INTIMA	3.72	19.09
TOTOCAL. LOTTO ALTRI	3.72	17.64
ABB.SPORTIVO E T.L.	3.86	16.76
FRANC. SPESE POSTALI	3.88	18.32
PIANTE E FIORI	3.91	17.20
CONIGLI	3.92	22.15
DENARO AI FIGLI	3.97	14.78
FARINA	4.10	22.51

segue Tavola 2 - Coefficiente di variazione della stima della spesa totale  
e percentuale di varianza spiegata dalla rarità dell'acquisto

Categorie di spesa	Coefficiente di variazione (x100)	Percentuale varianza dovuta alla rarità
ASS.AUTO E ALT. VEIC	4.16	23.15
DISCHI FONOCASSETTE	4.46	21.55
CINEMA TEATRO DISCOT	4.74	11.99
QUADERNI CANCELLERIA	4.75	15.93
SCOPE FIAMMIFERI	4.80	14.47
ABBON. RADIO TV	4.81	20.84
BEVANDE ANALCOLICHE	4.89	24.23
BIANCHERIA CASA	4.90	21.14
BIGL. ABB. AUTOBUS	5.19	17.89
BIRRA	5.22	25.94
INSETT.DISINF.VERNIC	5.33	25.70
OLII E ALTRI LUBRIF.	5.41	27.96
FRUTTA SECCA	5.43	24.29
SPESE ANIM. DOMEST.	5.73	26.25
LIBRI	5.73	15.60
GIOCATTOLI	5.85	18.93
CONFEZ.ARTICOLI ABB.	6.02	30.49
RISC.CENTRALE	6.13	42.65
MENSE	6.25	24.12
LIQUORI	6.61	22.26
OVINA E CAPRINA	6.67	30.02
CARBONE LEGNA	6.79	38.87
GAS BOMBOLE	6.90	32.89
ENEL II CASA	6.97	23.91
MANUT.ELETTRODOM.	6.99	28.90
GASOLIO ALTRI COMB.	7.01	20.26
ANALISI MEDICHE	7.15	29.97
CARNE IN SCATOLA	7.23	28.57
CARROZZERIA	7.41	19.08
FITTO FIGUR. II CASA	7.51	26.45
BORSE ARTICOLI PELLE	7.65	29.61

segue Tavola 2 - Coefficiente di variazione della stima della spesa totale  
e percentuale di varianza spiegata dalla rarità dell'acquisto

Categorie di spesa	Coefficiente di variazione (x100)	Percentuale varianza dovuta alla rarità
BIGLIETTI ABB. FERR.	7.93	31.46
ISTRUZIONE	7.95	24.46
ALTRE SPESE AUTO MOT	8.24	34.09
PNEUMATICI	8.33	33.40
SUCCHI DI FRUTTA	8.35	26.84
PICCOLI APP.ELETTTR.	8.36	34.29
LAMPADARI QUADRI	9.41	39.86
EQUINA	9.46	27.83
TESSUTI x CONF.MISU.	9.52	35.32
OCCHIALI SIRINGHE	9.58	45.80
ACQUISTO AUTO NUOVA	9.72	29.10
COPERTE TESS.ARRED.	9.73	32.38
LAVATRICI	9.74	24.10
ALTRI BENI CASA	9.97	33.36
SERVIZI DOMESTICI	10.45	35.45
TELEVISORI COLORI	10.51	31.64
KEROSENE GASOLIO	10.82	29.28
MEDICI INFERMIERI	11.10	23.50
GARAGE AUTO MOTO	11.14	32.07
ARTICOLI PERSONALI	11.16	43.80
STOVIGLIE PLASTICA	11.16	34.85
GAS BOMBOLE IICASA	11.25	46.22
SVILUPPO PELLICOLE	11.36	28.18
BICICLETTE	11.52	32.13
ARGENT.,GIOIELLI	11.58	53.12
FALEGNAME	11.95	36.72
MAN. MACC. FOT.E AL.	12.08	42.30
BRICOLAGE HOBBY	12.24	29.60
PIATTI E TAZZE	12.25	45.58
MATERASSI	12.37	44.97
FRUTTA CONSERVATA	12.38	27.10

segue Tavola 2 - Coefficiente di variazione della stima della spesa totale  
e percentuale di varianza spiegata dalla rarità dell'acquisto

Categorie di spesa	Coefficiente di variazione (x100)	Percentuale varianza dovuta alla rarità
ALTRE SPESE FINANZ.	12.83	28.22
RADIO RADIOGRAMMOFO.	12.86	42.29
TASSE SCOLASTICHE	13.07	31.26
ASSICUR. VITA	13.11	34.32
MOBILI LEGNO	13.29	35.55
CUCINE STUFE SCALD.	13.69	44.42
ACQUISTO AUTO USATA	13.96	33.92
VIDEO REGISTRATORI	14.45	36.29
IMPIANTI	14.62	31.84
MACCH.FOT.,CINEPRESE	14.71	41.06
FRIGORIFERI CONG.	14.87	35.32
ALTRE ASSICURAZIONI	14.93	35.76
MANUT.VESTIARIO	15.27	41.70
ACQUA II CASA	15.65	61.55
ALBERGHI PENSIONI	15.88	45.94
ALTRE SPESE RICREAT.	15.94	32.53
ARTICOLI SPOR. CAMP.	16.18	36.96
SP.CONDOM. II CASA	16.33	35.84
ONORARI LIB. PROFES.	16.34	41.88
LARDO	16.80	27.84
GIRADISHI REGISTRAT.	17.19	45.78
GITE VIAGGI TURIST.	18.54	34.41
PELLICCE E PELLE	19.19	44.14
ALTRE MEDICHE	19.72	58.03
LUCIDATRICI ASPIR.	19.78	54.05
PERSONAL COMPUTER	20.00	39.27
LAVASTOVIGLIE	20.33	29.32
MANUT.BIANCHERIA	20.54	38.86
CULTO	21.56	37.16
PENTOLE PADELLE POS.	21.61	26.11
TAPPETI MOQUETTE	21.92	39.06

segue Tavola 2 - Coefficiente di variazione della stima della spesa totale  
e percentuale di varianza spiegata dalla rarità dell'acquisto

Categorie di spesa	Coefficiente di variazione (x100)	Percentuale varianza dovuta alla rarità
ACQ. SCOOTERS MOTO	22.13	20.70
GAS RETE II CASA	25.15	42.55
CARBONE LEGNA II CASA	26.40	39.02
BICCHIERI BOTTIGLIE	26.60	31.21
STRUM. MUSICALI	27.04	36.89
ANIMALI DOMESTICI	27.54	46.28
MACCH. DA SCRIVERE	29.14	56.51
BIGL. MEZZI MARITI.	29.73	51.17
TELEVISORI BIANCONER	29.76	54.35
MACCH. CUCIRE	31.03	35.56
BIGLIETTI AEREI	33.28	62.93
AFFITTO II CASA	35.70	46.99
MOBILI METALLO	37.86	58.87
RISC. CENTRALE II CASA	37.92	18.18
ALTRE SPESE TRASPORT	40.27	33.26
KEROS. GASOLIO II CASA	41.43	63.29
CONDIZIONATORI	43.21	44.00
RICOVERI OSPEDALIERI	46.27	38.46
MOBILI PLASTICA	48.37	49.10
CASE RIPOSO	48.71	61.05
ACQUISTO ROULOTTE	53.77	34.35
BARCHE CANOTTI	61.77	47.99
IMPIANTI II CASA	68.92	53.04

Tavola 3 - Numerosità campionaria secondo la percentuale di acquirenti ipotizzata nell'universo

Frequenza acquirenti	Numerosità campionaria	Frequenza acquirenti	Numerosità campionaria
0.01	247500	0.45	3056
0.02	122500	0.50	2500
0.03	80833	0.55	2045
0.04	60000	0.60	1667
0.05	47500	0.65	1346
0.10	22500	0.70	1071
0.15	14167	0.75	833
0.20	10000	0.80	625
0.25	7500	0.85	441
0.30	5833	0.90	278
0.35	4643	0.95	132
0.40	3750	0.99	25

(\*) calcolata con riferimento ad uno schema di campionamento casuale semplice  
e con riferimento all'universo delle famiglie italiane nel 1989



Fig 1 - variabilita' della stima  
(spesa media pro-capite)

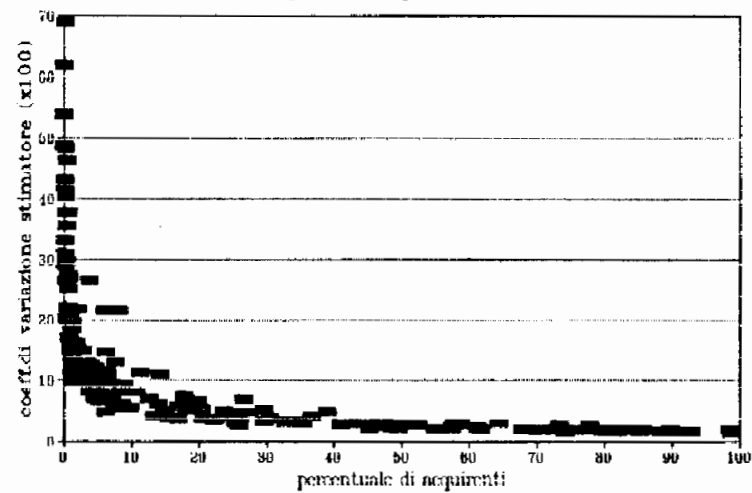


Fig 2 Contributo dei non acquirenti  
alla variabilita' - Valori percentuali

